Toward an Interim Standard for Patient-Centered Knowledge-Access*

Mark S. Tuttle[1], David D. Sherertz[1], Lawrence M. Fagan, PhD, MD,[2] Robert W. Carlson, MD,[2] William G. Cole, PhD,[3] Peter B. Schipma[4], Stuart J. Nelson, MD[5]
[1]Lexical Technology, Inc., Alameda, CA; [2]Stanford University, Stanford, CA ; [3]University of Washington, Seattle, WA; [4]I.S. Grupe, Lombard, IL; [5]Medical College of Georgia, Augusta, GA

Most care-giver "knowledge" needs arise at the point of care and are "patient-centered." Many of these knowledge needs can be met using existing on-line knowledge sources, but the process is too time-consuming, currently, for even the computer-proficient. We are developing a set of public domain standards aimed at bringing potentially relevant knowledge to the point of care in a straight-forward and timely fashion. The standards will a) make use of selected items from a Computer-based Patient Record (CPR), e.g., a diagnosis and measure of severity, b) anticipate certain care-giver knowledge needs, e.g., "therapy," "protocols," "complications," and c) try to satisfy those needs from available knowledge sources, e.g., knowledge-bases, citation databases, practice guidelines, and on-line textbooks. The standards will use templates, i.e., fill-in-the-blank structures, to anticipate knowledge needs and UMLS® Metathesaurus® enhancements to represent the content of knowledge sources. Together, the standards will form the specification for a "Knowledge-Server" (KS) designed to be accessed from any CPR system. Plans are in place to test an interim version of this specification in the context of medical oncology. We are accumulating anecdotal evidence that a KS operating in conjunction with a CPR is much more compelling to users than either a CPR or a KS operating alone.

> Details are all that matters: God dwells there, and you never get to see Him if you don't struggle to get them right. -- Stephen Jay Gould[1]

INTRODUCTION

Throughout the 1980's, observers noted that a) the time and place when care-givers need access to knowledge-sources is at the point-of-care, and b) access to knowledge sources could be improved if that access could take advantage of information already known to a Computer-based Patient Record (CPR). Since the '90s appears to be the decade of the CPR, it is time to plan for the smooth integration of knowledge-access into existing and future CPRs. We argue that the best way to achieve this integration is through the adoption of an interim standard for patient-centered knowledge-access. An "interim" standard would be devoted to a) what can be done now, and b) what it is practical to deploy. A long-term standard should follow the development of standards for a CPR.

Our argument has two parts: First, we will describe a simple method for incorporating selected pieces of information from a CPR into knowledge-source queries; second, we will propose that certain representations used to implement this method be adopted as the interim standard. The latter defines a way in which most knowledge-source schemas can be represented uniformly, and in which templates operating on the uniform representation can be used to anticipate care-giver knowledge needs. The ability to anticipate these needs in a simple but potentially reusable way is an important feature of the proposed standard. The workings of the proposed knowledge-access method will be illustrated using examples of some physician knowledge needs arising in the context of medical oncology. Some of these needs can be satisfied by accessing PDQ®, CANCERLIT®, and *Cancer: Principles & Practice of Oncology*[2] (P&PO), a textbook available in electronic form. Maximum use will be made of Unified Medical Language System®[3] (UMLS) content and methods.

Following emerging convention, we refer to data in a computer-based patient record as "information," and data in on-line books, citation databases and the like as "knowledge." The implication here is that the latter is both supported by an authoritative consensus and written at a level of abstraction intended to apply to more than one patient, e.g., "stage III rectal cancer."

THE PROBLEM

Patient encounters in a medical oncology clinic generate questions such as: "Should I do a bone scan?" "What are the appropriate staging studies?" "Are there any protocols?" "What's a tamoxifen flare?" "What should I do about an elevated alkaline phosphatase?" Using PDQ, CANCERLIT, and P&PO, a computer literate physician can find relevant information about each of these questions. The problem is that different skills are required to use each on-line source, and even a skilled user may require an hour or more to answer the questions generated by a single encounter.

Treating everything in these on-line sources as "text" to be searched using "words" doesn't solve the problem. All three sources contain the same words, but they have profoundly different content. For example, PDQ does not contain recommendations on staging studies; instead it provides detailed discussions regarding therapy *given the results of staging studies*. In principle, both the P&PO and CANCERLIT discuss the potential utility of a "bone scan" for, say, Stage I or Stage II breast cancer patients; but the 3rd (1989) Edition of P&PO says that "the value ... is a matter of controversy" (p. 1213) while CANCERLIT lists citations claiming that, for these patients, the true positive rate for the test is between 2% and 4%. Further, only P&PO contains instructions on

0195-4210/92/$5.00

how to plan staging studies given a diagnosis, and only PDQ contains information about standard and experimental protocols.

Lastly, a KS standard for oncology is unlikely to succeed unless at least portions of the standard are adopted for all of healthcare. Thus, part of the problem is developing an oncology knowledge-server in such a way that as many of its components and methods as possible can be reused by other healthcare specialties.

THE PROPOSED KNOWLEDGE-SERVER

A strategic goal of the National Cancer Institute (NCI) is to make oncology knowledge resources available at the point of care (R.J. Esterhay, MD, personal communication). Using speech, pen or keyboard input, a care-giver should be able to use a computer to consult PDQ, CANCERLIT or an on-line textbook, while in the process of caring for a patient. A key component of any such system will be a *knowledge-server* (KS); a software module that acts as an intermediary, or agent, handling the transactions between a care-giver's knowledge needs and the knowledge sources that are both available and potentially relevant to those needs.

We view the UMLS Metathesaurus as a "knowledge-server development enabling technology." Without it, a small inter-disciplinary team could not solve the knowledge access problem. The concept linking content and structures in the Metathesaurus need only be enhanced slightly to provide the connections we need in the domain of medical oncology. We call this enhanced version of the Metathesaurus "Meta+".

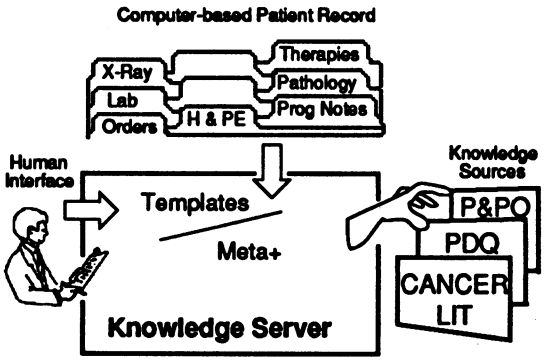

Figure 1 - A Knowledge-Server (KS) System Model: **Knowing the identity of the patient, the KS can try to fill the slots in its templates using information from a CPR. An enhanced version of the UMLS Metathesaurus (Meta+) is used to translate terms and retrieve knowledge (paragraphs) from the sources.**

Figure 1, above, attempts to answer the question, "What is a knowledge-server?" While this model has helped potential developers and users understand how a knowledge-server might work and why they might want one, it does not indicate how a knowledge-server might be deployed. Successful deployment is dependent upon software reusability and content extensibility, and these in turn are dependent upon *standards*. Software reusability means that components developed to make the knowledge-server work under one set of circumstances will be components that can be used again under other circumstances. Content extensibility means that new information (for example a new information source) can be added to the knowledge-server, without changing the components that are there already.

These details are important to Hospital Information System (HIS) / CPR vendors and their customers. Both groups need to be assured that a knowledge-server can be expanded, modified, maintained and tailored to local uses productive, and they want to see clearly where the boundaries are between their responsibilities and those of whoever owns and maintains the knowledge-server. Further they want to understand exactly how information is exchanged across these boundaries. Figure 2 shows the four places where standards will be necessary if software reusability and content extensibility are to be achieved. They are discussed below.

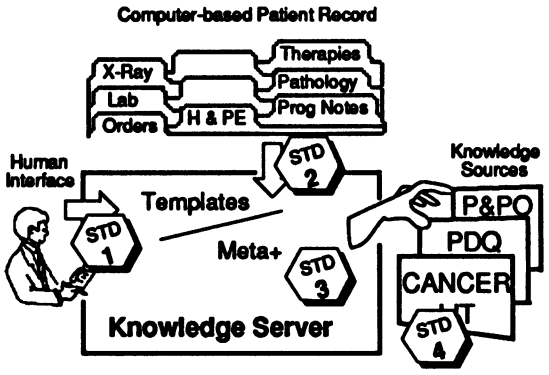

Figure 2 - The Interim Standards: **To be successful an interim standard ought to provide 1) a syntax and semantics for templates, 2) a way of asking for and receiving information from a CPR, 3) a way of enhancing the Metathesaurus, and 4) a way of representing sources.**

STD 1: A representation for templates, i.e., fill-in-the-blank structures (see Figure 3, below) that provide connections between users, patients, and sources of information.

STD 2: A specification for the information that will be exchanged between the knowledge server and the CPR, permitting the CPR to supply automatically what otherwise the care-giver would have to enter, manually.

STD 3: A representation for Metathesaurus enhancements so that the important aspects of the templates, knowledge sources and CPR not already covered by the Metathesaurus can be represented in a reusable way.

STD 4: A representation for knowledge sources, such as PDQ, CANCERLIT, or P&PO, so that a core set of KS procedures can be reused on each of these sources, and on new knowledge sources added later.

METHODS

To explore these issues using real knowledge needs and knowledge sources, we implemented a prototype[4] of a knowledge-server that supports access to P&PO, PDQ, and CANCERLIT. Early in our work on this prototype, we formulated three guiding questions: 1) What are some common physician knowledge needs generated by patient encounters? 2) What pieces of information from a CPR should be used to help satisfy those needs? 3) Exactly how should the information from the CPR be used by the knowledge-server to try to satisfy those needs?[5]

Our first step was to formulate two longitudinal scenarios, one for a breast cancer patient, and one for a rectal cancer patient. In both scenarios, new patient information tended to generate new information needs. One of us (RWC) attempted to meet these needs using the three on-line oncology knowledge sources, while recording the time required to use them. The scenarios provided specific answers to the questions above, e.g., "At this point in the scenario what information is likely to be in the patient's CPR?" Independently, the developer of Oncodisc® (PBS), and a medical librarian (Ms. Gloria Linder) each used the knowledge sources at their disposal to try to answer the same questions. LMF, DDS, and MST reviewed the results, and began to formulate some knowledge-server functions to expedite and improve the process.

RESULTS

Standard 1 - Templates

As shown in Figure 3, below, the *Master Template* is a central, organizing "point of view" for the oncology KS, implemented in the form of a frame with named slots and permissible values for each slot that depended on the values in other slots. The information in the right-hand column of this template is what allows other templates to begin to anticipate care-giver information needs. Care-givers see this "anticipation" in the form of, say, "buttons" on a screen that can acquire values from the *Master Template*.

Despite its simplicity, this organization proved to be both robust and very useful, even when the template is only partially filled in. For example, even if only the first four values (down through "Rectum") are known, this is enough for a "Print patient information" button in the knowledge-server interface to activate the *Patient Information Template* and retrieve paragraphs about "Rectal Cancer" from PDQ. Once a value for *Stage* is known, the same button can cause the same template to retrieve paragraphs regarding, say, "Stage III Rectal Cancer". Similarly, additional knowledge is retrievable after a value for *Histology* appears when the pathology report is available in the CPR. For example, the care-giver imperative "Protocols" will yield a shorter list of potentially relevant protocols than if it had been made prior to the availability of the pathology report because some "Rectal Cancer" protocols will be "ruled in" or "ruled out" given a value for *Histology* (*Adenocarcinoma*) and a *Stage* (*III*).

Slot Name	Slot Value
Patient Identifier	987654
Age	57
Sex	Male
Organ/System/Group	Rectum
Histology	Adeno-carcinoma
Stage	III
Previous Therapy	Resection
Current Therapy	CLB-9081

Figure 3 - Master Template for Rectal Cancer Scenario: **The values (right-hand column) have been filled in using information from a CPR.**

More specifically, the request "Staging Studies" will activate another template that retrieves the appropriate section of the P&PO chapter on "Colorectal Cancer". The same request "knows" to use "Rectal Neoplasms" and "Neoplasm Staging," both available in the Metathesaurus, with appropriate qualifiers, e.g., "radionucleotide imaging," as CANCERLIT search terms to retrieve more recent citations on the subject. Other anticipated requests, i.e., templates, include retrievals for "Therapy," "Complications," and the like.

Standard 2 - KS - CPR Information Exchange

Once the identity of the care-giver is known to the KS, the latter can start querying the CPR for new developments in the care-giver's patients. Given a patient identifier, basic queries like "What is the most recent cancer diagnosis available?", where the latter is specified by a range of ICD-CM-9 codes, will be part of the standard. Translating the query into something that the CPR can understand would be a local implementation task; similarly, the local implementation would have the CPR write any result to the KS in the prescribed form - say a code followed by a date. If the CPR had this information represented in another coding system, either Meta+ would have to be extended to "understand" that code, or the information would have to be presented on the screen for the care-giver to translate. The permissible (target) choices for the translation are part of the standard, i.e., the oncology KS is prepared to retrieve paragraphs about certain disease entities, called by one of these names, etc. If the CPR pathology report is not coded, the standard will provide a way to ask for and retrieve text from the CPR. The latter will be presented to the care-

giver to interpret, i.e., to convert to permissible slot values.

Standard 3 - Metathesaurus Enhancements
As described below under **Standard 4**, the KS assumes that every "formal" concept name in the templates and the knowledge sources appears in or has been added to the Metathesaurus. Maintaining the enhancements in the Metathesaurus syntax is easy; maintaining the Metathesaurus semantics is more difficult, though this process is gradually becoming exportable to sites outside the NLM (Betsy Humphreys, personal communication). In most cases, the desired concepts and names, e.g., "Breast Cancer", are already in the Metathesaurus and the only thing that needs to be added is the fact that some template or some source also uses that name to mean the same thing. Of course, the greater the overlap between any "formal" names in the CPR, e.g., the local names for lab tests, and the formal names in the templates and knowledge sources, the more useful the KS will be, but having this overlap is not a requirement. What is a requirement is that the enhancements "look" like any other Metathesaurus data so that procedures designed to manipulate them will not need to be changed.[6].

Standard 4 - Source Representation
Medical oncology covers only a relatively small number of "diseases," e.g., PDQ contains seventy-eight "Disease Entities", and the "Practice" part of P&PO contains thirty-eight chapters, one per cancer or cancer group, e.g., "Colo-rectal cancer." The potentially unbounded amount of detail present in the discussions of therapies and complications can all be "viewed" through the organization implied by the structure of the *Master Template*, because most oncology patients acquire an *Organ*, *Cell Type*, and *Stage* early in their diagnosis and treatment. It proved natural to reorganize the formal schemas of each of the three information sources to take advantage of this observation. In each case, the goal of the standard representation is to convert each source into paragraphs placed in a hierarchy. Each node in the hierarchy is a formal concept to be added to the Metathesaurus if it is not already there.

PDQ: Almost all the knowledge in PDQ, i.e., paragraphs in a database schema, can be reorganized into a hierarchy defined by its "Disease Entities". E.g.,

```
PDQ
  Gastrointestinal_Cancer
    Rectal_Cancer
      Cellular_Diagnosis_Rectal_Cancer
        Adenocarcinoma_of_the_rectum.
```

For every such "path" through PDQ there are "Allowable Questions", e.g.,

- Patient_Information
- Stage_Information
- Treatment_Options
- Protocols
- Supportive care
- Screening
- Prognosis

Somewhat less intuitively, almost all the information in the "Practice" part of P&PO and most of the citations in CANCERLIT can be organized into similar hierarchies.

CANCERLIT: Almost all the potentially relevant knowledge in CANCERLIT, i.e., citations with index terms, can be placed in two hierarchies - one classifying neoplasms by site, the other classifying them by histology. These hierarchies are the respective subtrees in MeSH (Medical Subject Headings), the naming system used to classify the citations. One source of "allowable questions" are the MeSH subheadings, e.g., "/complications".

P&PO: Almost all the knowledge in P&PO, i.e., the paragraphs in the chapters making up the "Practice" part of the textbook can be placed in a hierarchy extracted from the *Table of Contents*. One source of "allowable questions" are the chapter subheadings, e.g., *Staging*.

Browsing: When the templates fail to anticipate care-giver knowledge needs, the knowledge-server will need to support browsing. For example, we do not yet plan to make the *Complications Template* specific for answering the question, in the context of *Breast Cancer*, "What's a tamoxifen flare?" Fortunately, this is exactly the kind of thing an index is good for, and users who search CANCERLIT using the appropriate words will be led to some relevant papers on the subject. Indexes for all three sources will be prepared using the standard "word-index" software now available as part of the UMLS Knowledge Sources. Similarly, users wishing to browse "top down" can navigate using the hierarchy available for each source, or they can navigate using Meta+, as the latter will contain all the concepts and relationships from each knowledge source hierarchy. Having a Table of Contents, i.e., hierarchy, and Index for each source will not solve all browsing problems; but having them available uniformly, simply, and transparently, will make them easy to use, and care-givers will need to learn only a single set of "navigation" conventions.

DISCUSSION

The point of having a patient-centered knowledge-access standard is to create both intellectual and commercial economies of scale. If we could all access the same repertoire of national resources, and some local ones in addition, using the same, or similar, software, there would be more incentive to make knowledge sources available, and to maintain them, and more incentive to create and maintain the required software.

Generalizing to Other Specialties
While the prototype takes advantage of the fact that most cancer patients soon acquire a primary site, a histologic

classification, and, eventually, a stage, nothing we have done limits the applicability of the method to medical oncology. In fact, some early viewers of the prototype quickly proposed the development of templates for other specialties, e.g., pediatric pulmonology. One of us (SJN) will be examining the utility of representing MKSAP (Medical Knowledge Self-Assessment Program) and AHCPR Practice Guidelines in this way.

It will not be lost on most readers of this paper that a standard terminology for clinical medicine, nursing, and other domains, would eliminate about half of the work to be done to incorporate any new knowledge source. But we are still some distance from having a single such a corpus. In the near term, increasing clinical coverage in the Metathesaurus, e.g., that provided by SNOMED International (SNOMED III)[7], seems to be the best alternative available.

Near Term Goals

Our near term plans call for making a simple stand-alone version of the knowledge-server that collaborators and selected users may access over the Internet. This version will support three modes of access: UNIX-like commands (line-based), UNIX curses (ASCII terminal-based), and X-windows (bit-mapped-based).

We have tentative arrangements with five HIS/CPR vendors, namely, First Data Corp., HBO & Company, Second Foundation, Inc., SMS (Shared Medical Systems, Inc.), and TDS Healthcare Systems Corp., to evaluate the potential of a medical oncology knowledge-server accessible from their systems. Not surprisingly, the most difficult part of this evaluation will be the means by which information is extracted from the CPR. In some cases, the knowledge-server will emulate a user request for a patient attribute, and the CPR will write the result to a file where the knowledge-server will read it. Obviously, a longer-term objective is to develop some sort of standard encapsulation method that can be used by all vendors. Ongoing work at Columbia Presbyterian Hospital (Sideli, et al.) suggests that the HL/7 standard may suffice.

An Interim Standard

An interim standard for the KS should take into account users, patients, CPRs, and knowledge sources not only as they exist today but also as they will exist in the near future. A successful standard will provide incentives for the development of useful and compatible CPRs and KS; it may even help shape a consensus on design specifications for subsequent versions of these products.

ACKNOWLEDGMENTS

Robert J. Esterhay contributed several years of provocative discussions regarding the notion of a knowledge-server. Peter Walton, MD, President, Second Foundation, Inc., contributed his notion of a "plug and play" knowledge-server. John Foy, PhD, MD, and Charles Rogerson, et al., at TDS Healthcare Systems Corporation, helped clarify the issues implicit in attempts to support patient-centered knowledge-access in existing commercial CPRs. A dozen UMLS Contractor Meetings at the NLM provided twice-a-year confrontations with the difficulty of both defining and solving the patient-centered, knowledge-access problem. Employees at Lexical Technology, including Nels Olson, Mark Erlbaum, David Sperzel, Lloyd Fuller and Octavio Suarez-Munist, faculty, staff, and students at Stanford, including Keith Campbell and Diane Oliver, and several score collaborators and colleagues contributed directly or indirectly to the vision presented here. Finally, this work has benefited from similar ideas and technology emerging at the University of Pittsburgh (R. Miller, et al.), Columbia University (J. Cimino, et al.), and Beth Israel Hospital in Boston (C. Safran, et al.).

References

*Supported by NCI SBIR Contract #N43-CM-27759 and NLM Contract N001-LM-0-3515.

[1]. Gould, SJ, Eight Little Piggies: Reflections in Natural History, Norton, 1993, p. 14.

[2]. DeVita, VT, S Hellman, SA Rosenberg, editors, CANCER: Principles and Practice of Oncology, 3rd Edition, J. B. Lippincott, Philadelphia, 1989.

[3]Lindberg, DAB and BL Humphreys, "The UMLS Knowledge Sources: Tools for Better User Interfaces," R. Miller, editor, SCAMC, IEEE Computer Society Press, Los Alamitos, California, 1990, pp. 121-125.

[4]. Sherertz, DD, LM Fagan, MS Tuttle, RW Carlson, PB Schipma, WG Cole, SJ Nelson, "Speech-Driven, Patient-Centered Access to Three Oncology Knowledge Sources: How Might it Work?" (Theater Demonstration), to be presented, Safran, C., editor, SCAMC, 1993.

[5]. Tuttle, MS, DD Sherertz, LM Fagan, RW Carlson, WG Cole, SJ Nelson, "Using Patient Information to Simplify and Improve Access to Three Electronic Oncology Information Sources," (Abstract) M. Kahn, editor, AMIA Spring Congress, 1993.

[6]Tuttle, MS, D Sherertz, M Erlbaum, W Sperzel, L Fuller, N Olson, S Nelson, J Cimino, C Chute, "Adding Your Terms and Relationships to the UMLS Metathesaurus," M. Frisse, editor, SCAMC, 1991, pp. 219-223.

[7]. Cote, RA, DA Rothwell, JL Palotay, RS Beckett, editors, SNOMED International: The Systematized Nomenclature of Human and Veterinary Medicine, College of American Pathologists, Northfield, IL, 1993.